Necessary conditions for variational regularization schemes

Dirk Lorenz and Nadja Worliczek*

March 2, 2013

Abstract

We study variational regularization methods in a general framework, more precisely those methods that use a discrepancy and a regularization functional. While several sets of sufficient conditions are known to obtain a regularization method, we start with an investigation of the converse question: What are necessary conditions for a variational method to provide a regularization method? To this end, we formalize the notion of a variational scheme and compare three different instances of variational methods. Then we focus on the data space model and investigate the role and interplay of the topological structure, the convergence notion and the discrepancy functional. Especially, we deduce necessary conditions for the discrepancy functional to fulfill usual continuity assumptions.

AMS: 49N45, 54A10, 54A20
Keywords: Variational regularization, Tikhonov regularization, necessary conditions, general discrepancy term, Bregman distances

1 Introduction

By "variational regularization" we mean every method that is used to approximate an ill-posed problem by well-posed minimization problems. We start with a mapping $F : X \to Y$ between two sets X and Y and equations

$$Fx = y.$$

A common problem with inverse problems is that of instability, i.e. that arbitrary small disturbances in the right hand side y (e.g. by replacing a "correct" y in the range of F with one in an arbitrarily small neighborhood) may lead to unwanted effects such as that no solution exists anymore or that solutions with perturbed right hand side differ arbitrarily from the true solutions. In topological spaces X and Y we can formulate the problem of instability more precisely: The equation $Fx^{\text{exact}} = y^{\text{exact}}$ is unstable, if there exists a neighborhood $\mathcal{U}$ of x^{exact} such that for all neighborhoods $\mathcal{V}$ of y^{exact} there exists $y^\delta \in \mathcal{V}$ such that $F^{-1}(y^\delta) \cap \mathcal{U} = \emptyset$ (cf. [22]).

*Institute for Analysis and Algebra, TU Braunschweig, 38092 Braunschweig, Germany, d.lorenz@tu-braunschweig.de, n.worliczek@tu-braunschweig.de.

Variational regularization methods replace the equation $Fx = y$ by a minimization problem for an (extended) real valued functional such that the minimizers are suitable approximate solutions of the equation. The most widely used variational method is Tikhonov regularization [32], but other methods are used as well. Starting from a detailed analysis of this method in Hilbert spaces, there are several recent studies on Tikhonov regularization in the context of more general spaces like Banach spaces [19, 28] or even topological spaces [11, 12, 16, 27]. These works provide a quite general set of sufficient assumptions under which Tikhonov regularization has the desired regularizing properties, i.e. stable solvability of the minimization problems and suitable approximation of the true solution if the noise vanishes. These sufficient assumptions are helpful to check if a chosen setting for variational regularization is indeed suited. On the other hand, when designing a regularization method it would be helpful to know in advance which setting works and which is not going to work. Hence, in this paper we begin with a study of the converse analysis and aim at providing *necessary conditions* on variational methods such that regularization is achieved. Such conditions will also be helpful in designing new variational methods as they rule out several options. Moreover, necessary conditions are a further step towards the understanding of the nature of variational regularization.

We remark that we are aware that necessary conditions can not be expected to be very strong (as an example, a minimization problem can be changed quite arbitrary without changing the minimizer itself). However, there are already a few results of this flavor known in the context of the regularization of ill-posed problems which we list here:

Theorem 1.1 (No uniform bounded linear regularization, [10, Remark 3.5])**.** *If the linear and bounded operator $F : X \to Y$ between Hilbert spaces X and Y does not have closed range and $(L_\alpha)_{\alpha>0}$ is a family of linear and bounded operators from Y to X such that for all $x \in X$ is holds that $L_\alpha Fx$ converges to x for $\alpha \to 0$, then $(\|L_\alpha\|)$ is unbounded.*

In other words, linear regularization methods are necessarily not uniformly bounded.

The next example of a necessary condition deals with the problem of parameter choice. We need the Moore-Penrose pseudo-inverse $F^\dagger$ of a bounded linear mapping between Hilbert spaces, cf. [4].

Theorem 1.2 (Bakushinskii Veto, [2])**.** *Let $F : X \to Y$ be a bounded linear operator between Hilbert spaces and $(L_\alpha)_{\alpha>0}$ be a family of continuous mappings from Y to X. If there is a mapping $\alpha : Y \to]0, \infty[$ such that*

$$\limsup_{\delta\to 0}\{\|L_{\alpha(y^\delta)}y^\delta - F^\dagger y\| \ : \ y^\delta \in Y,\ \|y - y^\delta\| \leq \delta\} = 0$$

then $F^\dagger$ is bounded.

In other words, parameter choice rules which are valid in the worst-case setting and which work for ill-posed problems (i.e. unbounded $F^\dagger$) necessarily need to use the noise level.

An example for a-priori parameter choice rule was proven by Engl:

Theorem 1.3 (Decay conditions for a-priori parameter choice rules for linear methods, [10, Prop. 3.7] and [9])**.** *Let F and (L_α) be as in Theorem 1.1, and*

$\alpha :]0,\infty[\to]0,\infty[$ *be an a-priori parameter choice rule. Then it holds that*

$$\lim_{\delta \to 0} \sup\{\|L_{\alpha(\delta)} y^\delta - F^\dagger y \ : \ y^\delta \in Y, \ \|y - y^\delta\| \leq \delta\} = 0$$

if and only if

$$\lim_{\delta \to 0} \alpha(\delta) = 0 \quad \textit{and} \quad \lim_{\delta \to 0} \delta \|L_{\alpha(\delta)}\| = 0.$$

In other words, a-priori parameter choice rules necessarily need to fulfill certain decay conditions.

Finally we mention the "converse results" from [26] which say that for Tikhonov regularization in Hilbert spaces certain convergence rates imply that certain source conditions are fulfilled (see [13] for generalization to other regularization methods).

Before we start our investigation of necessary conditions for variational regularization in Section 3, we start with a section in which we formalize the notation of a "variational scheme" and investigate a few different variational methods.

2 Variational schemes: Tikhonov, Morozov, and Ivanov methods

In this section we formalize the notion of a variational scheme which can be used to build variational regularization methods. Basically, a variational scheme consists of all ingredients which are needed to classify and analyze the associated minimization problems and their minimizers under perturbations of the data y. Hence, it should encode information about the involved spaces and its notions of convergence and "proximity", the forward operator, and the objective functional to be minimized. However, we do not allow for totally arbitrary objective functionals but we rather use the intuition that a variational scheme involves two functionals: a "similarity measure" or "discrepancy functional" ρ and a "regularization functional" R. The functional ρ is used to measure "similarity" in the data space in the sense that $\rho(Fx, y)$ is small if x explains the data y well. The functional R on the solution space is used to measure how well x fits prior knowledge in the sense that $R(x)$ is small for an x which fulfills the prior knowledge well.

Definition 2.1 (Variational scheme)**.** By a *variational scheme* we understand a tuple $\mathcal{M} = ((X, \tau_X), (Y, \tau_Y, \mathcal{S}), \rho, R, F)$, consisting of

- topological spaces X and Y equipped with topologies τ_X and τ_Y, respectively, X is called *solution space* and Y is called *data space*,
- a sequential convergence structure $\mathcal{S}$ on Y.

  That is, $\mathcal{S}$ is mapping which maps any element in Y to a set of sequences in Y such that the constant sequence (y) is an element of $\mathcal{S}(y)$ and that if a sequence is in $\mathcal{S}(y)$ then so does any of its subsequences. Usually, we denote $(y_n)_n \in \mathcal{S}(y)$ by $y_n \overset{\mathcal{S}}{\to} y$ and say that (y_n) converges to y (with respect to $\mathcal{S}$), see also [3, §1.7].
- $\rho : Y \times Y \to [0, \infty]$ is the *discrepancy functional*, for which we assume that $\rho(y, y) = 0$ for all $y \in Y$,

- $R : X \to [0, \infty]$ is the *regularization functional*, and
- $F : X \to Y$ is a mapping or *forward operator*.

As in [17] we use topological spaces since the functionals we consider do no take any linear structure into account which would justify the use of linear or normed spaces. While the role of most other ingredients of a variational scheme is obvious, we remark on the sequential convergence structure $\mathcal{S}$: Often decaying noise is described in terms of norm-convergence, a notion which is not available or not even appropriate (see, e.g. [12] and [27]). Therefore, the sequential convergence structure will be used to describe "vanishing noise" in Y, i.e. the vanishing of noise is modelled by convergence of a sequence (y_n) to noise free data y w.r.t. $\mathcal{S}$. Note that we do not assume that convergence w.r.t. $\mathcal{S}$ is topological since this is not used in standard proofs for regularizing properties (e.g. [19]). Moreover, the topology τ_Y may induce a different convergence structure which is more tied to the mapping properties of F. In the case of a Banach space Y one can think of the following situation which is for example used in [19]: The sequential convergence structure is given by the convergence with respect to the norm in Y and τ_Y is the weak topology on Y. Of course, there will be further relations between τ_Y, $\mathcal{S}$ and ρ in the following, and indeed, Section 3 mainly deals with these relations, but for the general variational scheme we keep them mostly unrelated.

We mention that we included the value ∞ in the range of the discrepancy functional ρ and the regularization functional R to model that certain data may be considered "incomparable" or that certain solutions may be impossible. As usual, the value ∞ is excluded for minimizers by definition and we use the notation $\operatorname{dom} R = \{x \ : \ R(x) < \infty\}$ (similarly for ρ).

Variational regularization methods can be build from variational schemes as follows. Instead of solving $Fx = y$ we aim at two goals: Find an $x \in X$ such that

1. x explains the data y well, in the sense that $\rho(Fx, y)$ is small, and
2. x fits to our prior knowledge in the sense that $R(x)$ is small.

In other words: We have two objective functionals $x \mapsto \rho(Fx, y)$ and $x \mapsto R(x)$ which we would like to "jointly minimize". Of course, in general this is not possible, however, such problems go under the name of "multicriteria", "multiobjective" of "vector optimization". A core notion there is that of "Pareto-optimal solutions", i.e., solutions x^* such that there does not exists an x such that $R(x) \leq R(x^*)$ and $\rho(Fx, y) \leq \rho(Fx^*, y)$ and one of both inequalities is strict [6, §4.7]. Note that for "exact data", i.e. y^{exact} in the range of F, the notion of Pareto optimality induces a notion of generalized solutions of the equation $Fx = y$ (see [12] for a slightly different notion):

Definition 2.2. Let $((X, \tau_X), (Y, \tau_Y, \mathcal{S}), \rho, R, F)$ be a variational scheme and y^{exact} be in the range of Y. We say that $\bar{x}$ is a *ρ-generalized R-minimal solution* of $Fx = y^{\mathrm{exact}}$ if $\rho(F\bar{x}, y^{\mathrm{exact}}) = 0$ and $R(\bar{x}) = \min\{R(x) \ : \ \rho(Fx, y^{\mathrm{exact}}) = 0\}$.

Using the two objective functionals $\rho(F\cdot, y)$ and R we can build at least three different minimization problems which aim at finding Pareto optimal solutions. These three problems are well known in the inverse problems community and in fact can be traced back to the pioneering works in the Russian school:

- Tikhonov regularization [32]: For $\alpha > 0$ set $T_{\alpha,y}(x) := \rho(Fx, y) + \alpha R(x)$ and consider
$$T_{\alpha,y}(x) \to \min_{x \in X}. \tag{T_α}$$
In other words: Choose a weighting between "good data fit" and "good fit to prior knowledge" and minimize the weighted objective functional. In the context of multicriteria optimization this is known as *scalarization*.

- Ivanov regularization [20]: For $\tau > 0$ consider
$$\rho(Fx, y) \to \min_{x \in X} \quad \text{s.t.} \quad R(x) \le \tau. \tag{I_τ}$$
In other words: Choose the solution with the best data-fit which also fits the prior knowledge up to a predefined amount.

- Morozov regularization [25]: For $\delta > 0$ consider
$$R(x) \to \min_{x \in X} \quad \text{s.t.} \quad \rho(Fx, y) \le \delta. \tag{M_δ}$$
In other words: Choose the solution which fits best the prior knowledge among the ones which explain the data up to a predefined amount.

These methods are treated and compared e.g. in [21, Ch. 3.5] (where (I_τ) is called "method of quasi-solutions" and (M_δ) goes under the name "method of the residual") in the case of Banach spaces and $\rho(Fx, y) = \|Fx - y\|^p$ and $R(x) = \|Lx\|^q$ with a (possibly unbounded) linear operator L. Here we present results on the relation of the minimizers of these methods in our abstract framework of a variational scheme. Note that we do not pose any convexity assumptions on R or ρ.

Theorem 2.3. *Let $\mathcal{M}$ be a variational scheme according to Definition 2.1.*

1. *If there exists a unique solution x_τ of (I_τ) which fulfills $R(x_\tau) = \tau$, then it solves (M_δ) with $\delta = \rho(Fx_\tau, y)$.*

2. *If there exists a unique solution x_δ of (M_δ) which fulfills $\rho(Fx_\delta, y) = \delta$, then it solves ($\mathrm{I}_\tau$) with $\tau = R(x_\delta)$.*

3. *If there exists a unique solution x_α of (T_α) then it solves (I_τ) with $\tau = R(x_\alpha)$ and (M_δ) with $\delta = \rho(Fx_\alpha, y)$.*

Proof. 1. With $\delta = \rho(Fx_\tau, y)$, it is clear that x_τ is feasible for the optimization problem (M_δ) and the objective value is $R(x_\tau) = \tau$. Assume that there is a solution $\bar{x} \ne x_\tau$ of (M_δ) with $R(\bar{x}) \le \tau$. Then, $\bar{x}$ would be feasible for (I_τ) with objective $\rho(F\bar{x}, y) \le \delta = \rho(Fx_\tau, y)$ which is a contradiction to the uniqueness of the solution x_τ.

2. The proof mimics the proof of the first statement.

3. Again, assume that there exists a solution $\bar{x} \ne x_\alpha$ of (I_τ). Then, one sees that $T_{\alpha,y}(\bar{x}) \le T_{\alpha,y}(x_\alpha)$ contradicting the uniqueness of x_α. The proof is similar for the last claim.

□

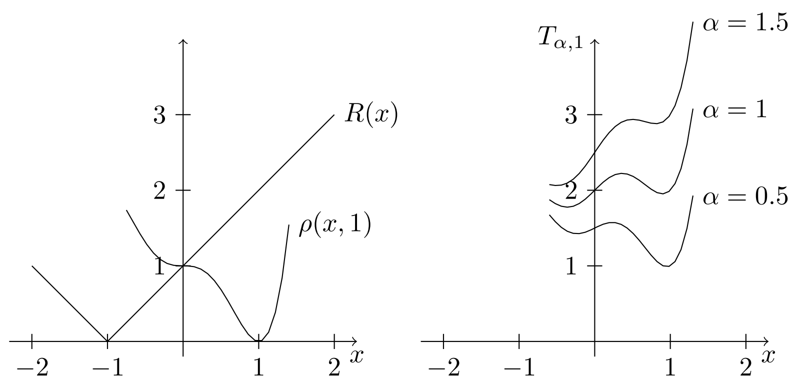

Figure 1: Illustration of Example 2.4

We remark that the missing implications in Theorem 2.3 are not true without additional assumptions.

Example 2.4 (Unique Ivanov and Morozov minimizers need not to be Tikhonov minimizers)*.* We illustrate this by a simple one-dimensional example: Let $X = Y = \mathbb{R}$, $F = \mathrm{id}$ and consider the regularization functional $R(x) = |x+1|$ (saying that the solution should be close to -1) and as discrepancy functional the so-called Bregman distance with respect to the strictly convex function $x \mapsto x^4$, i.e. $\rho(x,y) = y^4 - x^4 - 4x^3(y-x)$. We choose $\tau = 1$ and $y = 1$ and obtain $x_\tau = 0$ as the unique solution of (I_τ) (which is also the unique solution of (M_δ) with $\delta = 1$). But there is no $\alpha > 0$ such that $x = 0$ is a minimizer of $T_{\alpha,1}(x) = \rho(x,1) + \alpha|x+1|$ (cf. Figure 1).

In the above examples it holds that x_τ is a stationary point of the mapping $x \mapsto \rho(Fx,y)$. Moreover note that the precise form of R is not important in this example, several other R with $R'(0) > 0$ would also work. Indeed, we can deduce from the next proposition that it is necessary for x_τ to be also a (local) Tikhonov minimizer that not both of these properties are fulfilled.

Proposition 2.5. *Let $\mathcal{M}$ be a variational scheme and let X be a Banach space and let $y \in Y$. Furthermore, assume that the mappings $f(x) = \rho(Fx,y)$ and R are Gâteaux differentiable at $x^* \in X$.*

If x^* is a local minimizer of $T_{\alpha,y}$ for some $\alpha > 0$ then for every v it holds that

$$-\alpha R'(x^*)v \leq f'(x^*)v.$$

Moreover, if $R'(x^*) \neq 0$, then $f'(x^*) \neq 0$.

Proof. Since x^* is a local minimizer of $T_{\alpha,y}$, it holds, for $\epsilon > 0$ small enough,

$$\alpha(R(x^*) - R(x^* + \epsilon v)) \leq f(x^* + \epsilon v) - f(x^*).$$

Dividing by ϵ and passing to the limit $\epsilon \to 0$ proves the first assertion. If $R'(x^*) \neq 0$ then there is v such that $R'(x^*)v < 0$ and it follows that $f'(x^*)v > 0$; hence, $f'(x^*) \neq 0$. ☐

In other words: If we have a solution x^* of (I_τ) with $R'(x^*) \neq 0$ which is also a local minimizer of $T_{\alpha,y}$ then it is not stationary for $x \mapsto \rho(Fx,y)$.

Note that result similar to Proposition 2.5 can be found in [27, Thm. 4.13].

Remark 2.6. Under convexity assumptions on $f(x) = \rho(Fx, y)$ and R one can show that Ivanov minimizers (or Morozov minimizers) are indeed also Tikhonov minimizers for some parameter $\alpha > 0$ if they are not minimizers of the constraint. This is related to the fact that the subgradients of convex functions describe the normal vectors to the sublevel sets of the respective function, see e.g. [30].

Although the variational problems (T_α), (I_τ), and (M_δ) share their solutions under the circumstances presented above, they often differ with respect to their practical application.

It has been remarked already in early works (see, e.g., [31]) that Ivanov and Morozov regularization are related to different types of prior knowledge on the exact equation $Fx^{\text{exact}} = y^{\text{exact}}$. Morozov regularization is related to prior knowledge about the exact *data* or, more precisely, the *noise level* of the available data y, i.e., upper estimates on the quantity $\rho(y^{\text{exact}}, y)$. Ivanov regularization, however, is related to prior knowledge about the exact *solution*, more precisely, about upper estimates about the quantity $R(x^{\text{exact}})$.

Hence, the choice between Morozov and Ivanov regularization should be based upon the available prior knowledge at hand.

However, there are more factors, which should be taken into account when choosing the variational method, namely the factors of tractability and computational complexity. The three optimization problems (T_α), (I_τ), and (M_δ) may belong to different "subclasses" of optimization problems and their solution may have different computational complexity.

Example 2.7 (Linear problems in Hilbert space). In this classical setting, X and Y are Hilbert space, F is bounded and linear and we use $\rho(Fx, y) = \|Fx - y\|_Y^2$ and $R(x) = \|x\|_X^2$. In this case, the Tikhonov problem has an explicit solution $x_\alpha = (F^*F + \alpha\,\mathrm{id})^{-1}F^*y$ which can be treated numerically in several convenient ways (since the operator which has to be inverted is self adjoint and positive definite).

However, for both Ivanov and Morozov regularization no closed solution exists in general and one usually resorts to solving a series of Tikhonov problems, adjusting the parameter α such that the Ivanov or Morozov constraint is fulfilled [15].

Example 2.8 (Sparse regularization). We consider regularization of a linear operator equation $Ku = g$ with an operator $K : \ell^2 \to Y$ with a Hilbert space Y by means of a sparsity constraint [7, 18, 24]. In this setting one works with the discrepancy functional $\rho(Ku, g) = \frac{1}{2}\|Ku - g\|_H^2$ and the regularization functional $R(x) = \|u\|_1$ (extended by ∞ if the 1-norm does not exist). In this case:

- Tikhonov regularization consists of solving a convex, non-smooth, and unconstrained optimization problem (it is a non-smooth convex program, however, with additional structure),
- Morozov regularization consists of solving a non-smooth and convex optimization problem with a (smooth) convex constraint (and it can be cast as a second-order cone-program), and
- Ivanov regularization requires solving a smooth and convex optimization problem with a non-smooth convex constraint (it is a quadratic program).

Looking a little bit closer on this classification and the properties of ρ and R we observe that Ivanov regularization gives in fact the "easiest" problem since it obeys a smooth objective function and a constraint with a fairly easy structure (e.g. it is easy to calculate projections onto the constraint). On the other hand, the Morozov problem is "difficult" since it involves a non-smooth objective over a fairly complicated convex set (in the sense that projections onto the set $\|Ku - g\| \leq \delta$ are hard to calculate). Indeed, this rationale is behind the SPGL1 method [33, 34]: It replaces the Morzov problem with a sequence of Ivanov problems, solving each by a spectral projected gradient method, resulting in one of the fastest methods available for Morozov regularization.

It has been observed in different contexts, that the Ivanov approach yields faster algorithms than the Tikhonov approach in this case, e.g. the basic projected gradient method for Ivanov problems generally outperforms the basic iterative thresholding algorithm for Tikhonov problems [8].

In conclusion, the choice between the three variational methods should be based on the available prior knowledge and also on the tractability and the complexity of the corresponding optimization problem (often leading to a combination of two methods).

3 Necessary conditions for Tikhonov schemes

In this section we analyze regularization properties of the Tikhonov method. First we formalize our requirements for a scheme to be regularizing in the Tikhonov case. As usual we formulate conditions on existence, stability and convergence of the minimizers, cf. [29].

Definition 3.1 (Tikhonov regularization scheme)**.** A variational scheme $\mathcal{M}$ is called *Tikhonov regularization scheme*, if the following conditions are fulfilled:

(R1) Existence: For all $\alpha > 0$ and all $y \in Y$ it holds that $\operatorname{argmin}_{x\in X} T_{\alpha,y}(x) \neq \emptyset$.

(R2) Stability: Let $\alpha > 0$ be fixed, $y_n \overset{\mathcal{S}}{\to} y$ and $x_n \in \operatorname{argmin}_{x\in X} T_{\alpha,y_n}(x)$. Then (x_n) converges subsequentially and for each subsequential limit $\bar{x}$ of (x_n) it holds that $\bar{x} \in \operatorname{argmin}_{x\in X} T_{\alpha,y}(x)$.

(R3) Convergence: Let $Fx = y$ have an exact solution such that $R(x) < \infty$ and $y_n \overset{\mathcal{S}}{\to} y$. Then there exists a sequence $(\alpha_n)_n$ of positive real numbers such that $x_n \in \operatorname{argmin}_{x\in X} T_{\alpha_n,y_n}(x)$ converges subsequentially and every subsequential limit $\bar{x}$ is a ρ-generalized R-minimal solution of $Fx = y$.

3.1 Trivial necessary conditions

First we list fairly obvious necessary conditions to be regularizing in the Tikhonov sense. To that end, we introduce the solution operator

$$\begin{aligned} \mathcal{A} :& Y \times]0,\infty[\to 2^X \\ & (y,\alpha) \mapsto \operatorname*{argmin}_{x\in X} T_{\alpha,y}(x). \end{aligned}$$

for the Tikhonov problem (T_α). For fixed $\alpha > 0$ we denote $\mathcal{A}_\alpha(y) = \mathcal{A}(\alpha, y)$. We consider $\mathcal{A}$ and $\mathcal{A}_\alpha$ as set valued mappings and use the respective notation (see, e.g., [30]), especially the notion of the domain $\operatorname{dom}\mathcal{A}_\alpha = \{y \in Y \ : \ \mathcal{A}_\alpha(y) \neq \emptyset\}$ and the graph $\operatorname{gr}(\mathcal{A}_\alpha) = \{(y, x) \in Y \times X \ : \ x \in \mathcal{A}_\alpha(y)\}$.

Moreover, we recall that a topology is called *sequential* if it can be described by sequences, i.e. every sequentially closed set is closed.

Remark 3.2. Let $\mathcal{M}$ be a variational scheme. Then obviously (R1) is fulfilled if and only if $\operatorname{dom}\mathcal{A} = Y \times]0, \infty[$. This is the case if and only if for every y it holds that $\operatorname{dom} R \cap F^{-1}(\operatorname{dom}\rho(\cdot, y)) \neq \emptyset$, especially $\operatorname{range} F \cap \operatorname{dom}\rho(\cdot, y) \neq \emptyset$.

Theorem 3.3. *Let $\mathcal{M}$ be a variational scheme that fulfills (R2), $\alpha > 0$ and $y \in Y$. Then*

1. *$\mathcal{A}_\alpha(y)$ is sequentially compact and so is $\{x_n \ : \ x_n \in \mathcal{A}_\alpha(y_n)\} \cup \mathcal{A}_\alpha(y)$ for every sequence (y_n) in Y such that $y_n \xrightarrow{\mathcal{S}} y$.*

2. *The implication*

$$\left.\begin{array}{r} y_n \xrightarrow{\mathcal{S}} y \\ x_n \xrightarrow{\tau_X} x \\ x_n \in \mathcal{A}(\alpha, y_n) \end{array}\right\} \Rightarrow x \in \mathcal{A}(\alpha, y)$$

does hold, i.e. the mapping $\mathcal{A}_\alpha$ is sequentially closed.

If $\mathcal{S}$ is induced by a topology τ and $\tau_X \times \tau$ is sequential, then $\operatorname{gr}(\mathcal{A}_\alpha)$ is closed for every $\alpha > 0$.

If $\mathcal{A}_\alpha$ is single valued, then (R2) does hold if and only if $\mathcal{A}_\alpha$ is continuous w.r.t $\mathcal{S}$ and the sequential convergence structure of τ_X.

Proof. Let (x_n) be a sequence in $\mathcal{A}_\alpha(y)$ and consider the constant sequence $y_n := y$. Then $y_n \xrightarrow{\mathcal{S}} y$ and $x_n \in \mathcal{A}_\alpha(y_n)$ do hold. Therefore (R2) implies the existence of a convergent subsequence of (x_n) converging to an element of $\mathcal{A}_\alpha(y)$. □

3.2 A closer look on the data space

There exists a vast amount of settings that provide sufficient conditions for a Tikhonov scheme with non-metric discrepancy term to be regularizing. Here we start from a theorem which is extracted from [11, 12, 27].

Theorem 3.4. *Let $\mathcal{M} = ((X, \tau_X), (Y, \tau_Y, \mathcal{S}), \rho, R, F)$ be a variational scheme that fulfills the following list of assumptions:*

(A1) The sublevelsets $\{x \in X \ : \ R(x) \leq M\}$ are sequentially compact w.r.t τ_X for all $M > 0$ and R is sequentially lower semicontinuous

(A2) $\operatorname{dom} T_{\alpha,y} \neq \emptyset$ for all $y \in Y$

(A3) $(x, y) \mapsto \rho(Fx, y)$ is sequentially $\tau_X \times \tau_Y$ lower semi continuous

(A4) The sequential convergence structure $\mathcal{S}$ is given by

$y_n \xrightarrow{\mathcal{S}} y$ if and only if $\rho(y, y_n) \to 0$ *[CONV]*

and furthermore it fulfills

$y_n \xrightarrow{\mathcal{S}} y$ implies $\rho(z, y_n) \to \rho(z, y)$ for all $z \in \operatorname{dom}\rho(\,\cdot\,, y)$ *[CONT]*

(A5) $y_n \overset{\mathcal{S}}{\to} y$ *implies* $y_n \overset{\tau_y}{\to} y$

Then $\mathcal{M}$ is a Tikhonov regularization scheme.

Proof. We will only give a sketch of the proof, for details we refer to [12, 19, 27].

Since ρ and R are nonnegative (A1)–(A3) imply (R1) (existence of minimizers).

Let (x_n) be a sequence of minimizers as in (R2). Then $(R(x_n))$ is bounded due to (A2) and [CONT]. Hence, (A1) delivers a convergent subsequence. Let $\bar{x}$ be the limit of such a subsequence. Then, (A5), (A3) and [CONT] yield $T_{\alpha,y}(\bar{x}) \le T_{\alpha,y}(x)$ for all $x \in X$. Consequently, (R2) is fulfilled (stability).

Let $Fx^\dagger = y$, $R(x^\dagger) < \infty$ and (y_n) be a sequence such that $y_n \overset{\mathcal{S}}{\to} y$. Then, due to [CONV], there exists α_n such that

$$\alpha_n \to 0 \text{ and } \frac{\rho(y, y_n)}{\alpha_n} \to 0 \text{ as } n \to \infty \tag{1}$$

does hold (e.g. $\alpha_n = \sqrt{\rho(y, y_n)}$).

Therefore $R(x_n) \le \frac{1}{\alpha_n} T_{\alpha_n, y_n}(x^\dagger)$ for $x_n \in \operatorname{argmin}_{x \in X} T_{\alpha_n, y_n}(x)$ and together with (A1) this yields subsequential convergence and $R(\bar{x}) \le R(x^\dagger)$ for every subsequential limit $\bar{x}$. Using [CONV] we get $\rho(F(x_n), y_n) \to 0$, which yields $\rho(F\bar{x}, y) = 0$ due to (A5) and (A3). □

Remark 3.5. In [27] it is additionally assumed that $\rho(z, y) = 0$ implies $y = z$. This allows to formulate (R3) with R-minimal solutions in the strict sense (i.e. with $Fx = y$) instead of ρ-generalized R-minimal solutions.

In item (A4) it would be sufficient if [CONT] only holds for $z \in \operatorname{dom} \rho(\,\cdot\,, y) \cap F(X)$.

As remarked earlier, it is hard to obtain necessary conditions for a general Tikhonov scheme to be regularizing. Hence, we have chosen to start with the analysis of the data space Y. This is motivated by the fact that there are three different objects that pose additional structure on Y, namely the topology τ_Y, the sequential convergence structure $\mathcal{S}$ and the discrepancy functional ρ. Obviously, not every choice of these three objects will lead to a regularization scheme. We start from Theorem 3.4 and the conditions [CONV], [CONT] and (A5) and investigate the interplay of τ_Y, $\mathcal{S}$ and ρ and deduce necessary conditions on their relations. We are aware that the conditions [CONV], [CONT] and (A5) are not necessary for a scheme to be regularizing, but they appear as natural conditions in the context of regularization. However, we will get two different topologies that, under appropriate circumstances, provide exactly the desired convergent sequences, both given in a constructive way. Applied to specific classes of discrepancy functionals this could allow a deeper structural insight on what [CONT] does really mean and may tackle a subclass for which Theorem 3.4 is eligible without further adaptions.

Now we define the two topologies mentioned above, the first one designed to satisfy [CONV], the second to satisfy [CONT].

Definition 3.6. Let Y be a set and $\rho : Y \times Y \to [0, \infty]$ such that $\rho(y, y) = 0$ for all $y \in Y$.

1. We call
$$\mathcal{B}_\varepsilon^\rho(z) := \{y \in Y \ : \ \rho(z,y) < \varepsilon\}$$
the ε-ball w.r.t ρ centered at u and set
$$\tau_\rho := \{U \subseteq Y \ : \ \forall z \in U \, \exists \, \varepsilon > 0 \text{ such that } \mathcal{B}_\varepsilon^\rho(z) \subseteq U\}\,.$$

2. Let $Z, \tilde{Y} \subset Y$ and $[0,\infty]$ be equipped with the one-point compactification of the standard topology on $[0,\infty[$. For $z \in Z$ we define
$$f_z : \tilde{Y} \to [0,\infty] \text{ by } \tilde{y} \mapsto \rho(z,\tilde{y})\,.$$
By τ_{IN} we denote the initial topology on $\tilde{Y}$ w.r.t the family $(f_z)_{z\in Z}$ i.e. the coarsest topology on $\tilde{Y}$ for which all the f_z are continuous.

Note that the notation τ_{IN} does not reflect the dependency on $\tilde{Y}$ and Z. Hence, throughout the paper we will always mention explicitly the involved $\tilde{Y}$ and Z.

Remark 3.7. The two additional sets Z and $\tilde{Y}$ are introduced to allow to model a broader class of discrepancy functionals and to construct a larger variety of topologies. First, note that there are non-symmetric discrepancy functionals and even ones in which the domains of $\rho(\cdot,y)$ and $\rho(z,\cdot)$ differ. Especially, both arguments of ρ have different meanings: The first argument takes images of solutions x under F which can have additional structure (e.g. due to discretization), while the second argument takes measured data which may also have additional characteristics. Moreover, a smaller Z will allow for a coarser topology (and this will be helpful if the range of F is a "small" set) and a smaller $\tilde{Y}$ can model only a restrictive set of possible data (e.g. strictly non-negative one).

If $Z = \tilde{Y} = Y$ then τ_{IN} does satisfy [CONT]. Note that continuity of all the f_z on the whole of Y is stronger than required in [CONT]. There we only need continuity of f_z in Y for all $z \in \operatorname{dom}\rho(\,\cdot\,,y)$.

For the reader's convenience we recall some properties of the topologies τ_ρ and τ_{IN} that will be used in the further course of the paper.

Lemma 3.8. *The following properties hold for τ_ρ:*

1. *τ_ρ is a sequential topology.*

2. *A mapping from Y to an arbitrary topological space is τ_ρ-continuous if and only if it is sequentially continuous w.r.t τ_ρ.*

3. *$\rho(y,y_n) \to 0$ implies $y_n \overset{\tau_\rho}{\to} y$.*

The following holds for τ_{IN}:

4. *For arbitrary $Z, \tilde{Y} \subseteq Y$ sequential convergence w.r.t τ_{IN} can be characterized as follows:*

   Let $(y_n)_{n\in\mathbb{N}}$ be a sequence in $\tilde{Y}$ and $y \in \tilde{Y}$. Then $y_n \overset{\tau_{IN}}{\to} y$ if and only if $\rho(z,y_n) \to \rho(z,y)$ for all $z \in Z$.

5. *If additionally $\tilde{Y} \subseteq Z$ does hold, $y_n \overset{\tau_{IN}}{\to} y$ implies $\rho(y,y_n) \to 0$.*

Proof. For *1.* see [1, §2.4]. Item *2.* is a direct consequence of τ_ρ being sequential and *3.* is clear from the defintition open sets w.r.t. τ_ρ. Then, the first implication of *4.* is due to the sequential continuouity of continuous maps, the converse holds because the set $\{f_z^{-1}(V) \; : \; z \in Z,\, V \subseteq [0,\infty] \text{ open}\}$ is a subbase for τ_{IN}. Finally, *5.* is the continuity of f_y at y. □

Now we investigate the relation of τ_ρ to the property [CONV].

Theorem 3.9. *Let τ be a topology on Y. Then the following does hold:*

1. *The property*
$$\rho(y, y_n) \to 0 \text{ implies } y_n \xrightarrow{\tau} y \tag{2}$$
does hold if and only if τ is coarser than τ_ρ.

2. *If τ has property [CONV], then so does τ_ρ. In particular τ_ρ is the finest topology with that property.*

Proof. 1. Let τ be coarser than τ_ρ, then every τ_ρ-convergent sequence is also τ-convergent, and therefore (2) does hold.

Now let τ be a topology where (2) does hold. Suppose there exists $U \in \tau$ and $U \notin \tau_\rho$. Then there is an $u \in U$ such that for all $n \in \mathbb{N}$ there exists a $y_n \in \mathcal{B}_{\frac{1}{n}}(u) \setminus U$. Evidently $\rho(u, y_n) \to 0$ does hold and therefore $y_n \xrightarrow{\tau_Y} u$ in contradiction to $y_n \notin U$.

2. Let τ_Y be a topology that fulfills [CONV] and (y_n) a τ_ρ convergent sequence with limit y. Due to 1. τ is coarser than τ_ρ, therefore $y_n \xrightarrow{\tau_Y} y$ and consequently $\rho(y, y_n) \to 0$.

□

So, if $\mathcal{S}$ is induced by a topology at all, this is also done by the relatively well-behaved (i.e. sequential) topology τ_ρ. This is, e.g., the case if $\mathcal{S}$ provides unique limits since a sequence $\mathcal{S}$-converges given all its subsequences have a subsequence tending to the same limit (see e.g. [3, Prop. 1.7.15], [23]).

Corollary 3.10. *If there is a topology τ where $\rho(y, y_n) \to 0$ implies $y_n \xrightarrow{\tau} y$ such that [CONT] is fulfilled, then τ_ρ also fulfills [CONT].*

Since we are only interested in sequential convergence, this allows us to take τ_ρ as a sort of model topology.

Remark 3.11. In general, the set $\tau_{\mathcal{S}}$ of all sequentially open sets w.r.t to a sequential convergence structure $\mathcal{S}$ on Y is a topology on Y. As has been shown in [12, Prop. 2.10], in the case that $\mathcal{S}$ is given by [CONV], it is sufficient for [CONV] to hold for the topology $\tau_{\mathcal{S}}$ as well, that $\mathcal{S}$ fulfills [CONT].

Therfore assumption (A4) implies that τ_ρ also has [CONV] and this again implies that $\tau_{\mathcal{S}} = \tau_\rho$, since τ_ρ is sequential. Moreover, in this case the sets $\mathcal{B}_\varepsilon^\rho(y)$ are open for all $\varepsilon > 0$, $y \in Y$ (see also [12]) and therefore constitute a base for τ_ρ.

The next theorem deals with the question what consequences it has if [CONT] does hold in τ_ρ.

Theorem 3.12. *Let $Z \subseteq \bigcap_{y \in Y} \operatorname{dom} \rho(\,\cdot\,, y)$ be nonempty and $\tilde{Y} = Y$.*
If τ_ρ fulfills [CONT] then the following does hold:

1. *τ_{IN} is coarser than τ_ρ*

2. *If $Z = Y$ then τ_ρ and τ_{IN} both satisfy [CONV]. In particular they have the same convergent sequences.*

Proof.

1. Since $\rho(z, \cdot\,)$ is sequentially continuous for all $z \in Z$, it is also continuous and therefore τ_{IN} is coarser than τ_ρ.

2. Due to 1. convergence w.r.t. τ_ρ yields convergence w.r.t.. τ_{IN} and hence $\rho(y, y_n) \to 0$ implies $y_n \overset{\tau_{IN}}{\to} y$. Since $Y \subseteq Z$ the converse is also true and therefore τ_{IN} satisfies [CONV], and so does τ_ρ.

□

Remark 3.13. If $\tilde{Y} \subseteq Y$ and $(\tau_\rho)_{|\tilde{Y}}$ is sequential (e.g. if $\tilde{Y}$ open or closed w.r.t τ_ρ, see [14]), then $(\tau_\rho)_{|\tilde{Y}} = \tau_{\rho_{|\tilde{Y}}}$ does hold. In a setting where $\tilde{Y} \subsetneq Z \subseteq Y$, this together with Theorem 3.12 would still guarantee, that τ_{IN} and the subspace topology of τ_ρ on $\tilde{Y}$ provide the same convergent sequences.

If $\rho(z, \cdot\,)$ is τ_ρ-continuous at every $y \in Y$ for all $z \in Z$ regardless of the finiteness condition in [CONT], then we can drop the assumption $Z \subseteq \bigcap_{y \in Y} \operatorname{dom} \rho(\,\cdot\,, y)$ in Theorem 3.12.

So, in the setting of Theorem 3.12 sequential convergence in τ_ρ and τ_{IN} coincides. In general the sequential convergence structures of these topologies can be different from each other.

3.3 Application to Bregman discrepancies

We conclude Section 3 by an application to a special class of discrepancy functionals, namely ones that stem from Bregman distances which appear, e.g., in the case of Poisson noise or multiplicative noise [5]. Especially, this gives an example that illustrates how Theorem 3.12 can be used to gain necessary conditions on the discrepancy functional for Theorem 3.4 to apply. Also we treat the question, when $\rho(y_1, y_2) = 0$ implies $y_1 = y_2$ in this case.

In the following let V be a Banach space and $J : V \to [0, \infty]$ proper, convex, $Z = Y = \operatorname{dom} J$ and $\tilde{Y} \subseteq \{y \in Y \ : \ J$ Gâteaux differentiable at $y\}$. By ∇J we denote the Gâteaux derivative of J. As distance functional ρ we consider the Bregman distance w.r.t. to J. Consequently, for $(z, y) \in Y \times \tilde{Y}$ it takes the form

$$\rho(z, y) = J(z) - J(y) - \langle \nabla J(y), z - y \rangle.$$

The following lemma explores how convergence w.r.t τ_{IN} does actually look like and when $\rho(y_1, y_2) = 0$ does hold.

Lemma 3.14. *For all sequences (y_n) in $\tilde{Y}$, $y \in \tilde{Y}$ the following does hold:*

1. *$y_n \overset{\tau_{IN}}{\to} y$ if and only if $\rho(y, y_n) \to 0$ and $\langle \nabla J(y_n) - \nabla J(y), y - z \rangle \to 0$ for all $z \in Z$*

2. *If $\tilde{Y}$ is a linear subspace of V, then $y_n \overset{\tau_{IN}}{\to} y$ if and only if $\rho(y, y_n) \to 0$ and $\nabla J(y_n) \overset{*}{\rightharpoonup} \nabla J(y)$ in $\tilde{Y}^*$. In particular $(\nabla J)_{|\tilde{Y}} : \tilde{Y} \to \tilde{Y}^*$ is sequentially τ_{IN}-weak* continuous.*

3. *Let $y_1, y_2 \in \tilde{Y}$. Then $\rho(y_1, y_2) = 0$ if and only $\nabla J(y_1) = \nabla J(y_2)$.*

Proof. 1. The identity $\rho(z, y_n) - \rho(z, y) = \rho(y, y_n) + \langle \nabla J(y_n) - \nabla J(y), y - z\rangle$ does hold for all $z \in Z$.

So clearly $\rho(y, y_n) \to 0$ and $\langle \nabla J(y_n) - \nabla J(y), y - z\rangle \to 0$ imply $y_n \overset{\tau_{IN}}{\to} y$.

Conversely, let $y_n \overset{\tau_{IN}}{\to} y$ hold. Then $\rho(y, y_n) \to 0$ and hence $0 = \lim_{n\to\infty}(\rho(z, y_n) - \rho(z, y) - \rho(y, y_n)) = \lim_{n\to\infty}\langle \nabla J(y_n) - \nabla J(y), y - z\rangle$.

2. is a direct consequence of 1.

3. First let $\rho(y_1, y_2) = 0$. Then $J(y_1) = J(y_2) + \langle \nabla J(y_2), y_1 - y_2\rangle$ and hence linearity of $\nabla J(y_2)$ and nonnegativity of ρ imply $J(v) - J(y_1) - \langle \nabla J(y_2), v - y_1\rangle = \rho(v, y_2) \geq 0$ for all $v \in V$. Therefore $\nabla J(y_2)$ is a subgradient of J in y_1. Since J is differentiable at y_1 this yields $\nabla J(y_2) = \nabla J(y_1)$.

Now let $\nabla J(y_2) = \nabla J(y_1)$. Then $0 \geq -\rho(y_1, y_2) = \rho(y_2, y_1) \geq 0$.

□

Corollary 3.15. *Let $\operatorname{dom} J = V$ and J be differentiable on V and set $\tilde{Y} = V$.*

1. *If τ_ρ satisfies [CONT], then ∇J is τ_ρ-weak* continuous.*

2. *The property*

$$\rho(y_1, y_2) = 0 \Rightarrow y_1 = y_2 \text{ for all } y_1, y_2 \in Y$$

does hold if and only if ∇J is injective.

3. *τ_{IN} provides unique sequential limits if and only if ∇J is injective.*

So, in the setting of Corollary 3.15 it is necessary for Theorem 3.4 to apply to Bregman discrepancies that J has τ_ρ-weak* continuous derivative and a Tikhonov regularization scheme with discrepancy ρ guarantees convergence to an exact solution given J has injective derivative.

4 Conclusion

We examined variational regularization in a quite general setting and started a study on necessary conditions for variational schemes to be regularizing. Although it seems like little can be said about necessary conditions in general we obtained several results in this direction. Especially, we tried to clarify the relations between the different players in the data space, e.g. the convergence structure, the topology and the discrepancy functional. Here we started from a list of conditions which is known to guarantee regularizing properties and deduced necessary conditions for the topologies and the discrepancy functional. For Bregman discrepancies we illustrated that our results imply necessary conditions for the continuity of the derivative of the functional which induces the Bregman distance.

Although our results are fairly abstract, they are first steps towards the analysis of necessary conditions which can be used to figure out essential limitations of variational schemes. Next steps could be to analyze the other ingredients of a

variational scheme, namely the solution space X, its topology, the regularization functional and of course, the operator. Other directions for future research are to consider special classes of discrepancy functionals with additional structure and to extend the analysis to Morozov and Ivanov regularization.

References

[1] Alexander V. Arkhangel'skiĭ and Vitaly V. Fedorchuk. Basic concepts and constructions of general topology. In L. S. Pontryagin, editor, *General topology. I*, Encyclopaedia of Mathematical Sciences. Springer, 1990.

[2] Anatolii Borisovich Bakushinskiĭ. Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *Akademiya Nauk SSSR. Zhurnal Vychislitel′ noĭ Matematiki i Matematicheskoĭ Fiziki*, 24(8):1258–1259, 1984.

[3] Ronald Beattie and Heinz-Peter Butzmann. *Convergence structures and applications to functional analysis*. Kluwer Academic Publishers, Dordrecht, 2002.

[4] Adi Ben-Israel and Thomas N.E. Greville. *Generalized Inverses: Theory and Applications*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 15. Springer-Verlag, New York, second edition, 2003.

[5] Martin Benning and Martin Burger. Error estimates for general fidelities. *Electronic Transactions on Numerical Analysis*, 38:44–68, 2011.

[6] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004.

[7] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications in Pure and Applied Mathematics*, 57(11):1413–1457, 2004.

[8] Ingrid Daubechies, Massimo Fornasier, and Ignace Loris. Accelerated projected gradient method for linear inverse problems with sparsity constraints. *Journal of Fourier Analysis and Applications*, 14(5–6):764–792, 2008.

[9] Heinz W. Engl. Necessary and sufficient conditions for convergence of regularization methods for solving linear operator equations of the first kind. *Numerical Functional Analysis and Optimization*, 3(2):201–222, 1981.

[10] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of Inverse Problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 2000.

[11] Jens Flemming. Theory and examples of variational regularization with non-metric fitting functionals. *Journal of Inverse and Ill-Posed Problems*, 18(6):677–699, 2010.

[12] Jens Flemming. *Generalized Tikhonov regularization.* PhD thesis, Technische Universität Chemnitz, Chemnitz, may 2011.

[13] Jens Flemming, Bernd Hofmann, and Peter Mathé. Sharp converse results for the regularization error using distance functions. *Inverse Problems*, 27(2):025006, 18, 2011.

[14] Stanley P. Franklin. Spaces in which sequences suffice. *Fund. Math.*, 57:107–115, 1965.

[15] Andreas Frommer and Peter Maass. Fast CG-based methods for Tikhonov-Phillips regularization. *SIAM Journal on Scientific Computing*, 20(5):1831–1850 (electronic), 1999.

[16] Markus Grasmair. Multi-parameter Tikhonov regularisation in topological spaces. To appear in *Journal of Inverse and Ill-posed Problems*, 2011. http://arxiv.org/abs/1109.0364.

[17] Markus Grasmair. Multi-parameter tikhonov regularisation in topological spaces. arxiv/11090364, 2011.

[18] Markus Grasmair, Markus Haltmeier, and Otmar Scherzer. Sparse regularization with ℓ^q penalty term. *Inverse Problems*, 24(5):055020 (13pp), 2008.

[19] Bernd Hofmann, Barbara Kaltenbacher, Christiane Pöeschl, and Otmar Scherzer. A convergence rates result for Tikhonov regularization in Banach spaces with non-smooth operators. *Inverse Problems*, 23(3):987–1010, 2007.

[20] Valentin Konstantinovich Ivanov. On linear problems which are not well-posed. *Doklady Akademii Nauk SSSR*, 145:270–272, 1962.

[21] Valentin Konstantinovich Ivanov, Vladimir Vasilievich Vasin, and Vitaliĭ Pavlovich Tanana. *Theory of linear ill-posed problems and its applications.* Inverse and Ill-posed Problems Series. VSP, Utrecht, 2nd edition, 2002.

[22] Sergey I. Kabanikhin. *Inverse and Ill-Posed Problems: Theory and Applications.* De Gruyter, Berlin/Boston, 2011.

[23] Jan Kisyński. Convergence du type $\mathcal{L}$. *Colloquium Mathematicum*, 7:205–211, 1959/1960.

[24] Dirk A. Lorenz. Convergence rates and source conditions for Tikhonov regularization with sparsity constraints. *Journal of Inverse and Ill-Posed Problems*, 16(5):463–478, 2008.

[25] Vladimir A. Morozov. Choice of parameter for the solution of functional equations by the regularization method. *Doklady Akademii Nauk SSSR*, 175(6):1225–1228, 1967.

[26] Andreas Neubauer. On converse and saturation results for Tikhonov regularization of linear ill-posed problems. *SIAM Journal on Numeical Analysis*, 34(2):517–527, 1997.

[27] Christiane Pöschl. *Tikhonov Regularization with General Residual Term.* PhD thesis, Leopold Franzens Universität Innsbruck, Innsbruck, Austria, October 2008.

[28] Elena Resmerita and Otmar Scherzer. Error estimates for non-quadratic regularization and the relation to enhancement. *Inverse Problems*, 22(3):801–814, 2006.

[29] Otmar Scherzer, Markus Grasmair, Harald Grossauer, Markus Haltmeier, and Frank Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences.* Springer, New York, 2009.

[30] Winfried Schirotzek. *Nonsmooth Analysis.* Springer, Berlin, 2007.

[31] Vitaliĭ Pavlovich Tanana. The optimality of methods for the solution of unstable nonlinear problems. *Dokl. Akad. Nauk SSSR*, 220:1035–1037, 1975.

[32] Andreĭ Nikolaevich Tikhonov. Solution of incorrecly formulated problems and the regularization method. *Doklady Akademii Nauk SSSR*, 151:501–504, 1963.

[33] Ewout van den Berg and Michael P. Friedlander. SPGL1: A solver for large-scale sparse reconstruction, June 2007. http://www.cs.ubc.ca/labs/scl/spgl1.

[34] Ewout van den Berg and Michael P. Friedlander. Probing the Pareto frontier for basis pursuit solutions. *SIAM Journal on Scientific Computing*, 31(2):890–912, 2008.